DOCUMENT RESUME

ED 478 212 TM 035 079

AUTHOR	Doran, Harold C.
TITLE	Evaluating the Consequential Aspect of Validity on the Arizona Instrument To Measure Standards.
PUB DATE	2001-03-00
NOTE	10p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 21-25, 2003).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01 Plus Postage. PC Not Available from EDRS
DESCRIPTORS	*Academic Achievement; *Achievement Gains; Elementary Secondary Education; *High Stakes Tests; State Programs; Testing Programs; Validity
IDENTIFIERS	*Arizona; *Consequential Evaluation

ABSTRACT

The primary purpose of this investigation was to determine whether the positive consequences associated with high stakes test use and interpretation in Arizona were shared among all grade levels in the elementary school, not just the tested grades. In addition, a curriculum alignment variable was examined to observe its association with curricular and instructional change. The research design was quasi-experimental in nature using the posttest-only design with nonequivalent groups. The Arizona testing program tests were only administered in grades 3 and 5 at the elementary level. The nontested grades served as the comparison group, and the unit of analysis was the teacher. Kindergarten and grade 6 were excluded from the analysis as not all schools had both grades. The research sample was all teachers in grades 1 through 5 in a school district in Tucson, Arizona, representing about 15,000 students in all. Teachers answered a questionnaire about the testing program and curriculum. Findings provide evidence that the positive curricular consequences were not being shared equally between tested and nontested grades. Teachers in grades 1 and 2 reported significantly fewer changes in their academic curriculum than did teachers at grades 3, 4, and 5. Results provide sufficient evidence to question the consequential validity of the testing program in raising the quality of education. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 478 212

Evaluating the Consequential Aspect of Validity on the Arizona Instrument to Measure Standards

Harold C. Doran

University of Arizona

Paper Presented at the annual conference of the American Educational Research Association

New Orleans, Louisiana

2001

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

H. Doran

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM035079

BEST COPY AVAILABLE

High stakes tests have become a prominent tool in the systemic reform movement documenting the need for change and serving as the instrument of educational change. The Arizona Instrument to Measure Standards (AIMS) test was implemented as one component in the systemic reform movement with the intended purpose to raise the quality of education and to expand learning opportunities for all children in the State of Arizona. The test was not administered in all grades at the elementary, middle, or high school level. Instead, only students in Grades 3, 5, 8, and high school students participate in the testing program.

If the intent of the testing program was to raise the quality of education and expand learning opportunities for all children, then it would be expected that the positive consequences associated with test use and interpretation would be shared among all grades, not only the tested grades. Therefore the primary purpose of this investigation was to determine whether the positive consequences associated with high stakes test use and interpretation in Arizona were shared among all grade levels in the elementary school, not just the tested grades. Additionally, a curriculum alignment variable was examined to observe its association with curricular and instructional change.

A. Theoretical Framework

The role of educational testing has expanded in recent years with almost every educational reform movement relying on some form of assessment. Linn (1993) alleged that assessment is central to school reform for two reasons. First, they document the need for change. Results from criterion or norm-referenced tests provide evidence to indicate that public schools lack the quality desired from the public. Second, the assessment then

serves as the agent of the reform movement. The power of the assessment as the agent of reform lies in the high stakes that are often associated with test interpretation including rewards, sanctions, graduation, merit pay, or reconstitution.

Due to the centrality of assessments as instrument of educational reform, they continued to become commonplace in the systemic reform movement. Linn (2000) purported that assessments are popular instruments of reform for four reasons. First, they are relatively inexpensive to implement when compared to real and meaningful school change. Second, they can be externally mandated. Third, they can be rapidly implemented, and fourth the results are visible.

The promise of tests as a reform tool had been regarded as the darling of policymakers and the essential weapons system in educational disarmament. Additionally, tests had the specific and intended purpose of documenting the need for change and reforming classroom teaching.

This reliance upon tests to evaluate educational quality and as a policy instrument had placed a great deal of importance on the validity of an assessment to serve in such a capacity. The expanded role of test score use and interpretation to serve as a valid indicator of school reform reached outside the scope of traditional validity definitions and caused the educational measurement community to reexamine the role of validity theory.

Although the educational research community debated the inclusion of testing consequences as an aspect of validity theory, Messick's authoritative manuscript on validity in Educational Measurement (1989) and The Standards for Educational and Psychological Testing (1999) essentially codified the inclusion of the consequential aspect of validity. The Standards for Educational and Psychological Testing emphasized

the importance of testing consequences by including it in the validity and public policy sections.

The theoretical rationale for comparing the tested and nontested grades was provided by Lane et al. (1998). Lane et al. suggested

> It would be valuable to examine the differences in the nature and extent of the consequences of the assessment program for the tested and the nontested grades and for tested schools and nontested schools. If the intent of the assessment program is to improve student learning for all students regardless of whether they are in a tested grade or not, then the positive consequences should be shared among students in both tested and nontested grades. (p.27)

It was recognized that testing consequences could occur in numerous domains. Therefore, the opportunity to learn research (OTL) was selected to define the domain within which curricular change would occur. Three OTL variables were selected to assess the extent of curricular change attributable to the AIMS test.

B. Research Methods

The research design was quasi-experimental in nature using the Posttest-Only Design with Nonequivalent Groups. Although this design was *ex post facto* in nature, it permitted the researcher to investigate the evidence supporting the intended effects and provided direction for future research.

Selection of this design was chosen by necessity. The AIMS test was not administered in all grades, only Grades 3 and 5 in the elementary school. The nontested grades served as a comparison group. The use of a pretest measure was not feasible

given that the AIMS test was first administered in the spring of 2000. Therefore, any pretest measure intending to assess the curricular impacts of the AIMS test must have been used prior to any AIMS implementation. Although the use of a proxy measure as suggested by Cook and Campbell (1979) could be used which correlated with the posttest measure, no such proxy could be found in the review of literature.

The unit of analysis for this study was the teacher. Data on the curricular impact were collected from the tested grades (3 and 5) and the nontested grades (1, 2, and 4). Although two schools within the district included the sixth grade, most elementary schools did not. Therefore, the sixth grade has not been included in the analysis. Furthermore, kindergarten classes have been excluded from the analysis for two reasons. First, some schools had full day kindergarten classes while others did not. Second, kindergarten classes did not participate in any form of student assessment from the state or district. External assessments were more foreign to this grade level than any other in the elementary school.

The research sample consisted of all teachers in Grades 1 through 5 within one school district in Tucson, Arizona. The district was chosen due to its large and diverse student population. This district represented a total student population of approximately 15,000 students in the north-central part of Tucson.

The independent variable in this investigation (IV) was a discrete measure that had been dichotomized into tested grades (3 and 5) and nontested grades (1, 2 and 4). The dependent variable (DV) was a continuous measure describing the curricular and instructional impact. These data for the DV were collected from self-report surveys.

The data were first subjected to an exploratory factor analysis (EFA) using both principal components and varimax rotation. Both rotated and unrotated solutions were examined. Each factor within the questionnaire was analyzed to assess the internal consistency using coefficient alpha.

To examine the construct validity and dimensionality of the AIMS Questionnaire further, the data were examined using a Multitrait-Multimethod Method Matrix. The use of this matrix presented all intercorrelations of the data from a single administration of the survey instrument. Coefficient alpha was used as the reliability coefficient in the diagonals while the Pearson product-moment correlation was placed in the heterotrait-monomethod triangle. The Multitrait-Multimethod matrix is typically used to examine the correlations of the constructs with independent methods to assess convergent and discriminant validity. The use of this fragmented monomethod matrix was recognized as a limitation to the study of convergent and discriminant validity. However, the monomethod matrix permitted inferences regarding the validity of each dimension of the instrument through the examination of correlational patterns. Although the OTL variables were a subset of the same research construct, multidimensionality would be substantiated if the correlations among the three OTL variables were not too high.

Descriptive statistics including means and standard deviations were computed for preliminary analysis at the item level. Data were also presented using mean plots for each construct analyzed. Subscales were then created constituting the major variables of interest for the study (i.e., content coverage, content exposure, content emphasis, test preparation, general support, general understanding of test results, use of test results, and school reform). Descriptive statistics were computed for each subscale.

The primary research hypothesis was tested using the Dunn-Sidak planned comparison test. Additionally, a two-factor analysis of variance (ANOVA) was performed comparing the main effects among the grade levels, Title I schools, and the interaction effect. A post-hoc analysis was performed to analyze the data further.

The Pearson product-moment correlation coefficient was computed to examine the relationship between curriculum alignment and the instructional change variables (OTL and School Reform) and between the levels of support and the curricular impact factors to provide exploratory data on the relationships between teacher support for the assessment and classroom impact.

C. Data Sources/Instrumentation.

The AIMS Questionnaire was developed consisting of 50 items. The questionnaire was reviewed by the district's research proposal committee and superintendent to assess appropriateness. Classroom teachers within the district also reviewed the survey for clarity and appropriateness and to estimate the time needed to complete the survey.

The questions were developed based on substantial literature review reflecting the work and conclusions of other educational researchers conducting consequential validity studies specifically focusing on the impact of the assessment on classroom instruction. The survey was developed to assess eight dimensions. These included teacher support for the AIMS test, use of test results for curricular/instructional planning, general understanding of the test results, school reform, content coverage, content exposure, content emphasis, and test preparation.

D. Results

The results of the planned comparison assessing the differences in curricular impact between the tested and nontested grades was found to be statistically significant, accounting for 12% of the variance. The main effect analyzing the individual grade levels was found to be statistically significant, accounting for 22% of the variance, while the main effect of Title I status was not found to be statistically significant. The interaction between grade level and Title I status was not found to be statistically significant. The Tukey-Kramer test was performed as the post-hoc analysis for the omnibus test to further analyze the data. The post-hoc analysis revealed that the first grade mean was significantly lower than means for Grades 3, 4, and 5 but did not differ significantly from Grade 2. The Grade 2 mean was significantly lower than Grades 3, 4, and 5. The means among Grades 3, 4, and 5 did not differ significantly.

These results provided evidence to suggest that positive curricular consequences were not being equally shared between the tested and nontested grades. Further analysis revealed that Grade 4 teachers reported having curricular change similar to Grades 3 and 5. Grades 1 and 2 reported significantly fewer changes to their academic curriculum than did Grades 3, 4, and 5.

Rejection of the primary hypothesis provided sufficient evidence to question the consequential validity of the AIMS test to raise the quality of education and expand learning opportunities for all Arizona children. If expanded opportunities were significantly lower in nontested grades when compared to the tested grades, it was questionable whether the test was accomplishing its intended purpose. Therefore, validity evidence was lacking to support the intended effect.

E. Implications for educational policy

- Use matrix sampling techniques to encourage instructional change among all grades rather than testing the same grades each year.
- Implement a new form of the test each school year.
- Combine the results of the assessment with other critical indicators of educational success when making high-stakes educational decisions.
- Use assessment scores as longitudinal measures of individual student achievement rather than comparing schools.
- Only use tests for the intended purposes where sufficient validity evidence exists to support its claims.
- Encourage a frontloading curriculum alignment process instead of backloading the curriculum to the test.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

TM035079

I. DOCUMENT IDENTIFICATION:

Title: Evaluating the Consequential Aspect of Validity on the Arizona Instrument to Measure Standards	
Author(s): HAROLD C. DORAN	
Corporate Source:	Publication Date: 2001

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

☐

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

☒

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: Harold C. Doran	Printed Name/Position/Title: HAROLD DORAN Director of Research
Organization/Address: 675 N. Washington, Suite 220 Alexandria, VA 22314	Telephone: 703 647 1628 FAX: 703 647 1700
	E-Mail Address: hdoran@nasdc.org Date: 6/6/03

(Over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Lab, Bldg 075
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Lab, Bldg 075
College Park, MD 20742
Attn: Acquisitions